Assignment 9 Submission

CS 432

Timothy Filippone

2. Use cosine distance metric (chapter 8) not euclidean distance.

So you have to implement numpredict.cosine() instead of using

numpredict.euclidean() in:

https://github.com/arthur-e/Programming-Collective-Intelligence/blob/master/chapter8/numpredict.py

A cosine similarity function was implemented using this equation posted by Josh Graham on Slack.

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}}$$

Here is the actual implementation, it is based on `def getdistances.`

```
def cosine_sim(data,a):
    sum_magA = 0
    sum_magB = 0
    sumAB = 0
    for i in range(len(data)):
        b=data[i][0]
    for i in range(len(data)):
        sumAB += a[i]*b[i]
        sum_magA += pow(a[i],2)
        sum_magB += pow(b[i],2)
    sum_magA = sqrt(sum_magA)
    sum_magB = sqrt(sum_magB)
    return sumAB/(sum_magA*sum_magB)
```